Estimating multi-parameter partial Expected Value of Perfect Information from a probabilistic sensitivity analysis sample: a non-parametric regression approach. Supplementary online material.

Mark Strong PhD[a§], Jeremy E. Oakley PhD[b], Alan Brennan PhD[a]

a. School of Health and Related Research (ScHARR), University of Sheffield, 30 Regent Street, Sheffield S1 4DA, UK.

b. School of Mathematics and Statistics, University of Sheffield, Hicks Building, Hounsfield Road, Sheffield S3 7RH, UK.

[§] Corresponding author: email m.strong@sheffield.ac.uk; tel +44 (0)114 222 0812.

This is document contains the appendices for the paper *Estimating multi-parameter partial Expected Value of Perfect Information from a probabilistic sensitivity analysis sample: a non-parametric regression approach.*

Appendix A - Estimation of the standard error and upward bias of the partial EVPI

Standard error of the two-level Monte Carlo estimator

The two-level Monte Carlo estimator for the partial EVPI of parameters of interest $\mathbf{x}_i$ is typically written as

$$\frac{1}{K}\sum_{k=1}^{K}\max_d \frac{1}{J}\sum_{j=1}^{J}\mathrm{NB}(d,\mathbf{x}_i^{(k)},\mathbf{x}_{-i}^{(j,k)}) - \max_d \frac{1}{N}\sum_{n=1}^{N}\mathrm{NB}(d,\mathbf{x}^{(n)}), \tag{27}$$

where $\mathbf{x}_{-i}^{(j,k)}$ are samples drawn from the distribution of $\mathbf{X}_{-i}|\mathbf{X}_i = \mathbf{x}_i^{(k)}$. The number of inner loop samples is J, the number of outer loop samples is K, and N is the number of samples used to estimate the value of the baseline decision option.[9] If N is large, then the sampling variability in the partial EVPI estimator is dominated by the first term.

However, if we instead estimate the partial EVPI by

$$\begin{aligned} & \frac{1}{K}\sum_{k=1}^{K}\max_d \frac{1}{J}\sum_{j=1}^{J}\mathrm{NB}(d,\mathbf{x}_i^{(k)},\mathbf{x}_{-i}^{(j,k)}) - \frac{1}{K}\sum_{k=1}^{K}\frac{1}{J}\sum_{j=1}^{J}\mathrm{NB}(d^*,\mathbf{x}_i^{(k)},\mathbf{x}_{-i}^{(j,k)}) \qquad (28) \\ = & \frac{1}{K}\sum_{k=1}^{K}\left\{\max_d \frac{1}{J}\sum_{j=1}^{J}\mathrm{NB}(d,\mathbf{x}_i^{(k)},\mathbf{x}_{-i}^{(j,k)}) - \frac{1}{J}\sum_{j=1}^{J}\mathrm{NB}(d^*,\mathbf{x}_i^{(k)},\mathbf{x}_{-i}^{(j,k)})\right\}, \end{aligned}$$

where

$$d^* = \arg\max_d \frac{1}{K}\sum_{k=1}^{K}\frac{1}{J}\sum_{j=1}^{J}\mathrm{NB}(d, \mathbf{x}_i^{(k)}, \mathbf{x}_{-i}^{(j,k)}), \tag{29}$$

we can exploit the positive correlation between the two terms in equation (28). This results in a lower overall variance for the partial EVPI estimator, even for cases where N is chosen to be very large. The two-level Monte Carlo EVPI values reported in our two case studies were computed using this approach.

We can estimate the standard error of the partial EVPI computed via this latter approach as follows. We denote

$$l_1^{(k)} = \max_d \frac{1}{J}\sum_{j=1}^{J}\mathrm{NB}(d, \mathbf{x}_i^{(k)}, \mathbf{x}_{-i}^{(j,k)}) \tag{30}$$

and

$$l_2^{(k)} = \frac{1}{J}\sum_{j=1}^{J}\mathrm{NB}(d^*, \mathbf{x}_i^{(k)}, \mathbf{x}_{-i}^{(j,k)}). \tag{31}$$

Unless J is very small, the estimated standard error of the partial EVPI is given by

$$\sqrt{\frac{1}{K}\{\widehat{\mathrm{var}}(l_1) + \widehat{\mathrm{var}}(l_2) - 2\widehat{\mathrm{cov}}(l_1, l_2)\}}, \tag{32}$$

where $\widehat{\mathrm{var}}$ and $\widehat{\mathrm{cov}}$ are the usual sample estimators. For cases where J is very small (typically in the order of 10s), an extra term is required to account for the inner loop Monte Carlo variability. See Oakley et al (2010) for details of this and for the derivation of an estimator for the upward bias of the two-level Monte Carlo method.[9]

Upward bias of the two-level Monte Carlo estimator

This section follows a similar derivation given in Oakley *et. al.* (2010).[9]

Firstly, we denote the net benefit for outer sample k, averaged over the inner loop samples, as

$$\hat{\mu}_d^{(k)} = \sum_{j=1}^{J}\mathrm{NB}(d, \mathbf{x}_i^{(k)}, \mathbf{x}_{-i}^{(j,k)}), \tag{33}$$

and the vector of $\hat{\mu}_d^{(k)}$ over decision options as $\hat{\boldsymbol{\mu}}^{(k)} = (\hat{\mu}_1^{(k)}, \dots, \hat{\mu}_D^{(k)})^{\mathrm{T}}$.

Unless J is trivially small the sampling distribution for $\hat{\boldsymbol{\mu}}^{(k)}$ has approximately a $D-$dimensional multivariate Normal distribution

$$N_D\left(\boldsymbol{\mu}^{(k)}, \frac{1}{J}\boldsymbol{V}^{(k)}\right), \tag{34}$$

where $\boldsymbol{\mu}^{(k)}$ and $\boldsymbol{V}^{(k)}$ are unknown.

Next, we define a variance-covariance matrix $\hat{\boldsymbol{V}}^{(k)}$ where element p, q of $\hat{\boldsymbol{V}}^{(k)}$ has value

$$\hat{V}_{p,q}^{(k)} = \frac{1}{J-1}\sum_{j=1}^{J}\left\{\mathrm{NB}(p, \mathbf{x}_i^{(k)}, \mathbf{x}_{-i}^{(j,k)}) - \hat{\mu}_{p,k}\right\}\left\{\mathrm{NB}(q, \mathbf{x}_i^{(k)}, \mathbf{x}_{-i}^{(j,k)}) - \hat{\mu}_{q,k}\right\}. \tag{35}$$

If we approximate $\boldsymbol{\mu}^{(k)}$ by $\hat{\boldsymbol{\mu}}^{(k)}$ and $\boldsymbol{V}^{(k)}$ by $\hat{\boldsymbol{V}}^{(k)}$ we can generate samples $\tilde{\boldsymbol{\mu}}_1^{(k)}, \ldots, \tilde{\boldsymbol{\mu}}_S^{(k)}$ from $N_D\left(\hat{\boldsymbol{\mu}}^{(k)}, \frac{1}{J}\hat{\boldsymbol{V}}^{(k)}\right)$ and hence observe the properties of the two-level Monte Carlo estimator for partial EVPI.

To estimate the upward bias we generate $\tilde{\boldsymbol{\mu}}_1^{(k)}, \ldots, \tilde{\boldsymbol{\mu}}_S^{(k)}$ for large S (say 10,000) from the distribution above, with

$$\tilde{\boldsymbol{\mu}}_s^{(k)} = \left(\tilde{\mu}_{1,s}^{(k)}, \ldots, \tilde{\mu}_{D,s}^{(k)}\right)^{\mathrm{T}}. \tag{36}$$

An estimate of the upward bias for outer sample k is given by

$$\hat{b}^{(k)} = \frac{1}{S}\sum_{s=1}^{S} \max\left\{\tilde{\mu}_{1,s}^{(k)}, \ldots, \tilde{\mu}_{D,s}^{(k)}\right\} - \max\left\{\hat{\mu}_1^{(k)}, \ldots, \hat{\mu}_D^{(k)}\right\}, \tag{37}$$

and the overall expected bias is

$$\hat{b} = \frac{1}{K}\sum_{k=1}^{K} \hat{b}^{(k)}. \tag{38}$$

Standard error and upward bias of GAM estimator

We can estimate the standard error and upward bias of the partial EVPI obtained by the GAM regression method using the following sampling approach.

Any GAM model can be re-expressed as a parametric model. All this requires is that we find the matrix $\boldsymbol{X}^*$ that maps the model coefficients $\hat{\beta}_d$ onto the fitted values $\hat{\mathbf{g}}_d = \{\hat{g}(d, \mathbf{x}_i^{(1)}), \ldots, \hat{g}(d, \mathbf{x}_i^{(N)})\}$, i.e.

$$\hat{\mathbf{g}}_d = \boldsymbol{X}_d^*\hat{\beta}_d. \tag{39}$$

Helpfully, $\boldsymbol{X}_d^*$ is returned by the `predict.gam` function in the `mgcv` package. Given $\boldsymbol{X}_d^*$ and $\boldsymbol{V}_{\beta_d}$, the covariance matrix for $\hat{\beta}_d$ (which is returned as part of the `gam` function call), then the estimated covariance for $\mathbf{g}_d|\mathbf{y}_d$ is

$$\hat{\mathbf{V}}_d = \boldsymbol{X}_d^* V_{\beta_d} \boldsymbol{X}_d^{*\mathrm{T}}. \tag{40}$$

The joint distribution of $\hat{\beta}_d$ is multivariate Normal, and therefore

$$\mathbf{g}_d|\mathbf{y}_d \sim N(\hat{\mathbf{g}}_d, \hat{\mathbf{V}}_d). \tag{41}$$

For each decision option d, we draw a large number (say 10,000) of sampled values of $\mathbf{g}_d|\mathbf{y}_d$ from the above distribution. We denote these samples $\tilde{\mathbf{g}}_d^{(s)}$, $(s = 1, \ldots, S)$. For each $\tilde{\mathbf{g}}_d^{(s)}$ we calculate the partial EVPI via equation (9) replacing $\hat{\mathbf{g}}_d$ with $\tilde{\mathbf{g}}_d^{(s)}$. We denote the sampled partial EVPI values $\tilde{e}_s$, $(s = 1, \ldots, S)$. The sample standard deviation of $\tilde{e}_s$ is an estimate of the standard error we require.

An estimate of the upward bias of the partial EVPI estimator due to the maximisation in equation (9) is given by

$$\hat{b} = \frac{1}{S}\sum_{s=1}^{S} \tilde{e}_s - \hat{e}. \tag{42}$$

where $\hat{e}$ is the partial EVPI estimate computed at $\hat{\mathbf{g}}_d$.

Standard error and upward bias of the Gaussian process estimator

We can estimate the standard error of the partial EVPI obtained by the Gaussian process method using the same sampling approach as above. The conditional distribution of $\mathbf{g}_d = \{g(d, \mathbf{x}_i^{(1)}), \dots, g(d, \mathbf{x}_i^{(N)})\}$ given the net benefits $\mathbf{nb}_d = \{\mathrm{NB}(d, \mathbf{x}^{(1)}), \dots \mathrm{NB}(d, \mathbf{x}^{(N)})\}$ is approximately multivariate Normal

$$\mathbf{g}_d | \mathbf{y}_d \sim N(\hat{\mathbf{g}}_d, \hat{\mathbf{V}}_d), \tag{43}$$

where $\hat{\mathbf{g}}_d$ is given in equation (21), and the estimated covariance matrix $\hat{\mathbf{V}}_d$ is

$$\hat{\mathbf{V}}_d = \widehat{\sigma}_d^2 \{\Sigma_d - \Sigma_d \Sigma_d^{*-1} \Sigma_d + (H - \Sigma_d \Sigma_d^{*-1} H)(H^{\mathrm{T}} \Sigma_d^{*-1} H)^{-1} (H - \Sigma_d \Sigma_d^{*-1} H)^{\mathrm{T}}\}. \tag{44}$$

The standard error and upward bias are estimated in the same manner as for the GAM method, as explained above.

Appendix B - Estimation of Gaussian process hyperparameters

For each decision option, d, we wish to find values for the hyperparameters $\boldsymbol{\delta} = (\delta_1, \dots, \delta_p)$ and ν that maximise the log posterior density $\pi(\boldsymbol{\delta}, \nu | \mathbf{nb}_d)$. Up to some additive constant, the log posterior density of $\boldsymbol{\delta}$ and ν given the net benefits $\mathbf{nb}_d$ is

$$\begin{aligned} \pi(\boldsymbol{\delta}, \nu | \mathbf{nb}_d) \quad = \quad & -\frac{n - q + 2a}{2} \log \left\{ \frac{(n - q - 2)\widehat{\sigma}^2 + 2b}{2} \right\} \\ & -\frac{1}{2} \log |\Sigma^*| - \frac{1}{2} \log |H^{\mathrm{T}} \Sigma^{*-1} H| + \pi(\boldsymbol{\delta}, \nu), \end{aligned} \tag{45}$$

where Σ is given by equation (17) and $\Sigma^* = \Sigma + \nu I$. The term $\widehat{\sigma}^2$ is the posterior mean for σ^2 given by equation (20), and a and b are the parameters of an Inverse Gamma prior density for σ^2. The final term $\pi(\boldsymbol{\delta}, \nu)$ is the joint prior density for $\boldsymbol{\delta}$ and ν. The derivation of equation (45) is given in appendix C.

For the correlation lengths δ_j we assume weak Normal priors $\log(\delta_j) \sim N(0, 10^6)$. For the variance and nugget terms we assume Inverse Gamma priors $\sigma^2 \sim IG(0.001, 0.001)$ and $\nu \sim IG(0.001, 1)$.

The log posterior equation (45) must be maximised numerically. Methods include deterministic algorithms such as Nelder-Mead, or stochastic algorithms such as simulated annealing. R code for the optimisation of the log posterior is available at:

`http://www.shef.ac.uk/scharr/sections/ph/staff/profiles/mark.`

Appendix C - Derivation of the posterior density of the GP regression hyperparameters

The likelihood of the net benefits $\mathbf{nb}_d$ under the Gaussian process model in equation (18), as a function of the hyperparameters $\boldsymbol{\beta}$, σ^2, δ_j and ν, is given by

$$l\{(\beta, \sigma^2, \boldsymbol{\delta}, \nu); \mathbf{nb}_d\} = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}} |\Sigma^*|^{\frac{1}{2}}} \exp\left\{ -\frac{1}{2\sigma^2} (\mathbf{nb}_d - H\boldsymbol{\beta})^{\mathrm{T}} \Sigma^{*-1} (\mathbf{nb}_d - H\boldsymbol{\beta}) \right\}, \quad (46)$$

where $\Sigma^* = \Sigma + \nu I$, and where Σ is the function of δ_j given by equation (17). Given a non-informative prior for $\boldsymbol{\beta}$, $\pi(\boldsymbol{\beta}) \propto 1$, and some arbitrary prior $\pi(\sigma^2, \boldsymbol{\delta}, \nu)$, where σ^2, $\boldsymbol{\delta}$ and ν are independent of $\boldsymbol{\beta}$ then the posterior density of $\boldsymbol{\beta}$, σ^2, δ_j and ν is

$$p(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\delta}, \nu | \mathbf{nb}_d) \propto \frac{\pi(\sigma^2, \boldsymbol{\delta}, \nu)}{(\sigma^2)^{\frac{n}{2}} |\Sigma^*|^{\frac{1}{2}}} \exp\left\{ -\frac{1}{2\sigma^2} (\mathbf{nb}_d - H\boldsymbol{\beta})^{\mathrm{T}} \Sigma^{*-1} (\mathbf{nb}_d - H\boldsymbol{\beta}) \right\}. \quad (47)$$

We define $\widehat{\boldsymbol{\beta}}$ to be the posterior mean for $\boldsymbol{\beta}$ as given in equation (19) (derivation not shown). By combining equation (47) with equation (19) we can re-express equation (47) in the form of a Normal Inverse Gamma density, allowing us to integrate out $\boldsymbol{\beta}$ giving

$$p(\sigma^2, \boldsymbol{\delta}, \nu | \mathbf{nb}_d) \propto \frac{\pi(\sigma^2, \boldsymbol{\delta}, \nu)}{(\sigma^2)^{\frac{n-q}{2}} |\Sigma^*|^{\frac{1}{2}} |H^{\mathrm{T}} \Sigma^* H|^{\frac{1}{2}}} \exp\left\{ -\frac{1}{2\sigma^2} (\mathbf{nb}_d - H\widehat{\boldsymbol{\beta}})^{\mathrm{T}} \Sigma^{*-1} (\mathbf{nb}_d - H\widehat{\boldsymbol{\beta}}) \right\}. \quad (48)$$

Next we define $\widehat{\sigma}^2$ to be the posterior mean of σ^2 given in equation (20) (derivation not shown), and re-express Eg. (48) in terms of $\widehat{\sigma}^2$ to give

$$p(\sigma^2, \boldsymbol{\delta}, \nu | \mathbf{nb}_d) \propto \frac{\pi(\sigma^2, \boldsymbol{\delta}, \nu)}{(\sigma^2)^{\frac{n-q}{2}} |\Sigma^*|^{\frac{1}{2}} |H^{\mathrm{T}} \Sigma^* H|^{\frac{1}{2}}} \exp\left\{ -\frac{(n-q-2)\widehat{\sigma}^2}{2\sigma^2} \right\}. \quad (49)$$

If we choose as a prior for σ_d^2 an Inverse Gamma $IG(a, b)$ density, then we can re-express equation (48) as

$$p(\sigma^2, \boldsymbol{\delta}, \nu | \mathbf{nb}_d) \propto \frac{\pi(\boldsymbol{\delta}, \nu_d)}{(\sigma^2)^{\frac{n-q+2a+2}{2}} |\Sigma^*|^{\frac{1}{2}} |H^{\mathrm{T}} \Sigma^* H|^{\frac{1}{2}}} \exp\left\{ -\frac{(n-q-2)\widehat{\sigma}^2 + 2b}{2\sigma^2} \right\}. \quad (50)$$

This posterior is also proportional to an Inverse Gamma density, which allows us to integrate out σ^2 to give

$$p(\boldsymbol{\delta}, \nu | \mathbf{nb}_d) \propto \frac{\pi(\boldsymbol{\delta}, \nu)}{|\Sigma^*|^{\frac{1}{2}} |H^{\mathrm{T}} \Sigma^* H|^{\frac{1}{2}}} \left\{ \frac{(n-q-2)\widehat{\sigma}^2 + 2b}{2} \right\}^{-\frac{(n-q+2a)}{2}}. \quad (51)$$

Taking the log gives equation (45).